DOCUMENT RESUME

ED 410 268 TM 027 057

AUTHOR Zeng, Lingjia
TITLE An IRT Scale Transformation Method for Parameters Calibrated from Multiple Samples of Subjects. ACT Research Report Series 96-2.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa.
PUB DATE May 96
NOTE 19p.
AVAILABLE FROM ACT Research Report Series, P.O. Box 168, Iowa City, IA 52243-0168.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Estimation (Mathematics); *Item Response Theory; Mathematics Tests; *Scaling; *Test Items
IDENTIFIERS Calibration; Item Parameters

ABSTRACT

A problem frequently confronted in item response theory (IRT) applications is that the item parameters calibrated using more than two independent samples of subjects must be expressed on the same scale. The existing methods were developed for a pairwise transformation, that is, from one scale to the other. The purpose of this study is to introduce a common scale transformation method that can simultaneously find a vector of transformation functions for placing the parameter estimated from two or more item pools on a specified common scale. Two examples are presented to illustrate the usefulness of the method. The first is from 2 forms of a mathematics test administered to nonequivalent groups of examinees (1,637 and 1,636 examinees). The second used the item parameter estimates of the 11 common items for the first item pool of the first study. The proposed common scale transformation method provides an alternative approach for solving the problem of incompatible IRT-parameter estimates calibrated with two or more independent samples of subjects. (Contains 2 tables and 11 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ACT Research Report Series 96-2

ED 410 268

An IRT Scale Transformation Method for Parameters Calibrated From Multiple Samples of Subjects

Lingjia Zeng

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM027057

BEST COPY AVAILABLE

ACT

May 1996

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243-0168

©1996 by The American College Testing Program. All rights reserved.

An IRT Scale Transformation Method For Parameters Calibrated From Multiple Samples of Subjects

Lingjia Zeng

Abstract

A problem frequently confronted in IRT applications is that the item parameters calibrated using more than two independent samples of subjects must be expressed on the same scale. The existing methods were developed for a pairwise transformation, that is, from one scale to the other. The purpose of the present study is to introduce a common scale transformation method which can simultaneously find a vector of transformation functions for placing the parameter estimates from two or more item pools on a specified common scale. Two examples are presented to illustrate the usefulness of the method.

Acknowledgment

The author thanks Bradley Hanson, Tianyou Wang and Michael Kolen for their comments and suggestions in improving the manuscript.

An IRT Scale Transformation Method For Parameters Calibrated From Multiple Samples of Subjects

Because of the indeterminant nature of the latent variable IRT models, the parameter estimates obtained from different independent calibrations may not be on the same scale. A linear transformation can be performed to place them on an arbitrary scale while preserving the same item characteristic functions. The scale transformation procedure requires that at least some of the items be common to the different calibrations. A number of procedures have been proposed for IRT scale transformation. Marco (1977) proposed using the mean and standard deviation of the *b*-parameter estimates of the common items to determine the transformation function. Loyd and Hoover (1980) described a method which used the mean of the b-parameter estimates and mean of the a-parameters estimates. Haebara (1980) and Stocking and Lord (1983) introduced loss-minimization methods for computing the transformation coefficients. These methods are based on minimizing a loss function that reflects the errors involved in the process of transforming the estimates of parameters in one metric to another. Baker and Al-Karni (1991) compared a loss-minimization method (Stocking & Lord, 1983) with a summary statistics based method (Loyd & Hoover, 1980). They found that the loss-minimization method was less sensitive to atypical combinations of underlying ability, item difficulty, and discrimination than are the methods based on summary statistics.

IRT parameter scale transformation is a directional process, that is, from one scale to another scale. Baker and Al-Karni (1991) used the terms *from* and *to* to make the direction of transformation explicit. The existing methods are all for a pairwise transformation, that is, there is one from scale and one to scale. If there is a need to compare parameter estimates from more than two item pools, the current practice is to find the transformation functions in a number of independent pairwise processes. The purpose of the present study is to introduce a common scale transformation method which can find a vector of transformation functions that place parameter estimates from two or more item pools on a common scale in one minimization process. Two examples are presented to illustrate the usefulness of the method.

Method

According to the three-parameter logistic model, $P_i(\theta, a_i, b_i, c_i)$, the probability of a correct response to item i by a person with ability level θ, is defined as:

$$P_i = c_i + \frac{1 - c_i}{1 + e^{-1.7a_i(\theta - b_i)}}, \tag{1}$$

where a_i is usually called the item discrimination, b_i the item difficulty, and c_i the probability that a person with a very low ability gives a correct response. By the nature of this logistic model, the origin and unit of the ability and difficulty metric are undetermined. A linear relationship exists between any pair of IRT parameter scales (Lord, 1980).

Transformation for Two Item Pools

Consider two item pools: Pool 1 and Pool 2, which share a set of k common items. Pool 1 and Pool 2 were calibrated using two nonequivalent random samples drawn from Population 1 and Population 2, respectively. Let $\hat{\omega}_{1i}$ denote a vector of item parameter estimates: $\hat{a}_{1i}$, $\hat{b}_{1i}$ and $\hat{c}_{1i}$ for common item i in Pool 1, and $\hat{\omega}_{2i}$ denote a vector of item parameter estimates: $\hat{a}_{2i}$, $\hat{b}_{2i}$ and $\hat{c}_{2i}$ for common item i in Pool 2. Because $\hat{\omega}_{1i}$ and $\hat{\omega}_{2i}$ are on different parameter scales, $P(\theta, \hat{\omega}_{1i})$ and $P(\theta, \hat{\omega}_{2i})$ are not necessarily the same even though they are for the same item.

Suppose $\hat{\omega}_{2i}$ is transformed via some linear transformation to $\hat{\omega}_{(1,2)i}$, that is on the parameter scale defined in calibrating Pool 1, such that

$$\hat{a}_{(1,2)i} = \frac{\hat{a}_{2i}}{A_{(1,2)}}, \quad \hat{b}_{(1,2)i} = A_{(1,2)}\hat{b}_{2i} + B_{(1,2)}, \text{ and } \hat{c}_{(1,2)i} = \hat{c}_{2i}, \tag{2}$$

where $A_{(1,2)}$ and $B_{(1,2)}$ are the slope and intercept of the linear transformation function. After these transformations, $\hat{\omega}_{1i}$ and $\hat{\omega}_{(1,2)i}$ are on the same parameter scale. Let θ_1 denote an ability level on the scale defined in calibrating Pool 1, then the following relationship can be established:

$$P(\theta_1, \hat{\omega}_{1i}) = P(\theta_1, \hat{\omega}_{(1,2)i}). \tag{3}$$

Suppose the abilities of the subjects in Population 1 has a known probability density function (pdf) $g(\theta_1)$. The expected proportion correct (EPC) score for item i given $\hat{\omega}_{1i}$ for the potential subjects from Population 1 is

$$S(\hat{\omega}_{1i}) = \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{1i}) g(\theta_1) d\theta_1, \tag{4}$$

and the EPC score for item i given $\hat{\omega}_{(1,2)i}$ for the same population of subjects is

$$S(\hat{\omega}_{(1,2)i}) = \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{(1,2)i}) g(\theta_1) d\theta_1, \tag{5}$$

From Equation 3 we know that Equations 4 and 5 are approximately equal, that is

$$S(\hat{\omega}_{1i}) \approx S(\hat{\omega}_{(1,2)i}), \tag{6}$$

or

$$\int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{1i}) g(\theta_1) d\theta_1 \approx \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{(1,2)i}) g(\theta_1) d\theta_1 \tag{7}$$

For one item and the identical c-parameter estimates, we can find a transformation that will make the left and right sides of Equations 6 and 7 exactly the same. For two or more items, there may not exist a transformation that will make the left and right sides of Equations 6 and 7 exactly the same for all items. The difference is

$$\delta_i = S(\hat{\omega}_{1i}) - S(\hat{\omega}_{(1,2)i}) = \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{1i}) g(\theta_1) d\theta_1 - \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{(1,2)i}) g(\theta_1) d\theta_1. \tag{8}$$

By minimizing δ_i for all the k common items, the common scale transformation coefficients $A_{(1,2)}$ and $B_{(1,2)}$ can be found. Thus, a loss function can be defined as

$$Q(A_{(1,2)}, B_{(1,2)}) = \sum_{i=1}^{k} \delta_i^2. \tag{9}$$

A set of transformation coefficients which minimize the loss function Q can be found by solving the following equation system:

$$\begin{aligned} \frac{\partial Q(A_{(1,2)}, B_{(1,2)})}{\partial A_{(1,2)}} &= 0, \\ \frac{\partial Q(A_{(1,2)}, B_{(1,2)})}{\partial B_{(1,2)}} &= 0, \end{aligned} \tag{10}$$

The partial derivatives needed in Equations System 10 are derived as follows.

$$\frac{\partial Q(A_{(1,2)}, B_{(1,2)})}{\partial A_{(1,2)}} = -2 \sum_{i=1}^{k} \delta_i \frac{\partial S(\bar{\omega}_{(1,2)i})}{\partial A_{(1,2)}}, \tag{11}$$

where

$$\frac{\partial S(\bar{\omega}_{(1,2)i})}{\partial A_{(1,2)}} = \int_{-\infty}^{\infty} \frac{-1.7(1-c_{2i})a_{2i}(\theta_1 - B_{(1,2)})e^{-1.7a_{2i}(\theta_1 - B_{(1,2)} - A_{(1,2)}b_{2i})A_{(1,2)}^{-1}}}{A_{(1,2)}^2[1+e^{-1.7a_{2i}(\theta_1 - B_{(1,2)} - A_{(1,2)}b_{2i})A_{(1,2)}^{-1}}]^2} g(\theta_1)d\theta_1. \tag{12}$$

$$\frac{\partial Q(A_{(1,2)}, B_{(1,2)})}{\partial B_{(1,2)}} = -2 \sum_{i=1}^{k} \delta_i \frac{\partial S(\bar{\omega}_{(1,2)i})}{\partial B_{(1,2)}}, \tag{13}$$

where

$$\frac{\partial S(\bar{\omega}_{(1,2)i})}{\partial B_{(1,2)}} = \int_{-\infty}^{\infty} \frac{-1.7(1-c_{2i})a_{2i}e^{-1.7a_{2i}(\theta_1 - B_{(1,2)} - A_{(1,2)}b_{2i})A_{(1,2)}^{-1}}}{A_{(1,2)}[1+e^{-1.7a_{2i}(\theta_1 - B_{(1,2)} - A_{(1,2)}b_{2i})A_{(1,2)}^{-1}}]^2} g(\theta_1)d\theta_1. \tag{14}$$

The integrals involved in the partial derivatives can be approximated numerically to any specified degree of precision using a Gauss quadrature formula. The pdf g can be estimated from

the calibration samples. If the marginal maximum likelihood estimation method(e.g., BILOG, Mislevy & Bock, 1991) is used to calibrate the item parameters, an empirical θ distribution is estimated along the item parameters. The empirical θ distribution consists of a set of quadrature points and associated weights which can be used in the loss function.

Alternatively, any distribution that is on the θ scale of the Pool 1 items can be used for computing the loss function, because Equations 3, 6 and 7 hold for any θ value. A simple choice is a uniform distribution. Using a uniform θ distribution to compute the expected proportion correct score is analogous to using an arbitrary set of points along the θ scale to compute the true score. This procedure is used in an implementation of the Stocking & Lord method (Baker & Al-Karni, 1991).

The loss function defined in Equation 9 is different from the loss function in the Stocking & Lord method. The major difference is that Stocking & Lord's loss function is based on the squared difference between the two true scores computed with the two sets of parameter estimates for the common items whereas the loss function introduced in the present study is based on the squared differences between two EPC scores summed over the individual common items.

The loss functions of the proposed method and Stocking & Lord's method differ from that of Haebara's method in two aspects. First, Haebara's loss function consists of two components, one is based on the discrepancies resulted in transforming scale 2 to scale 1 and the other is based on the discrepancies resulted in transforming scale 1 to scale 2 using the inverse of the scale 2 to scale 1 transformation function. The loss functions of the proposed and Stocking & Lord's methods do not contain the second component in Haebara's loss function. Second, in Haebara's loss function, the discrepancies between two sets of item parameter estimates are squared at various individual ability levels for all the individual common items before the summation over items. An investigation of the advantages and disadvantages for these loss functions is beyond the scope of the present paper but certainly deserves further study.

Using the EPC scores has some potential advantages in developing an iterative procedure for finding a common scale for the item pools that involve differential item functioning (DIF) items.

Some authors (e.g. Kim & Cohen, 1992, Lautenschlager & Park, 1988) pointed out that the scale transformation may be seriously affected by the presence of DIF items. After an iteration of scale transformation, the EPC scores can be computed for all common items. If the EPC scores differ too much for a common item, then this item may be suspected as a DIF item. The next iteration of scale transformation can performed without the suspicious DIF item and the EPC scores are compared again to detect additional suspicious DIF items. The iteration procedure stops when no more suspicious DIF items are involved in the common item set. Detailed technical and practical issues will be addressed in another study.

Transformation for Multiple Item Pools

The method developed for scale transformation for two item pools can be extended for three or more item pools. First consider an example of 3 item pools. Pool 1 and Pool 2 share a set of items that are common to both, and Pool 2 and Pool 3 share a set of items that are common to both. Pool 1 and Pool 2 have a direct link because they have a common item set. Pool 1 and Pool 3 do not have a direct link because they do not have a common item set, but they have an indirect link because both pools have a direct link to Pool 2. If we want to compare items in Pool 1 with items in Pool 3, using pairwise transformation methods, we need to first transform parameters of items in Pool 1 to the scale of Pool 2 then use the Pool 2 to Pool 3 transformation function to further transform them to Pool 3 scale. Alternatively, Pool 2 can be transformed to Pool 1, and Pool 3 can be transformed to Pool 2, then to Pool 1. In common scale transformation method, a vector of transformation functions that transform the parameters estimates of all the from scales to the to scale are found in one minimization process.

In the general case, m (m>2) item pools are calibrated using m independent samples of subjects. In order to perform a common scale transformation, we assume that every pair of the m item pools are linked either directly by a common set of items, or indirectly through a chain of pools which are directly linked pairwisely. Let us assume that the scale of Pool 1 items is referenced as a common scale, so the parameter estimates of the other m-1 item pools are to be transformed to this scale. The selection of the common scale is arbitrary. Any of the m scales can be referenced as a

common scale. Let $\hat{\omega}_{(1,j)i}$ denote a transformation for the parameter estimate of common item i in Pool j to scale 1, that is

$$\hat{a}_{(1,j)i} = \frac{\hat{a}_{ji}}{A_{(1,j)}}, \quad \hat{b}_{(1,j)i} = A_{(1,j)}\hat{b}_{ji} + B_{(1,j)}, \text{ and } \hat{c}_{(1,j)i} = \hat{c}_{ji}, \tag{15}$$

where $A_{(1,j)}$ and $B_{(1,j)}$ are the slope and intercept of the linear transformation function that transforms the parameter estimates from scale j to scale 1 and j=2 to m.

Let ϕ denote a set of indices for all the items that are common to at least two item pools. Let ψ_i denote a set of indices for the pools that contain item i. For example, $\psi_i = \{1,2,3\}$ means that item i is a common item shared by Pools 1, 2 and 3. After all the parameters are transformed to scale 1, the expected proportion correct scores for item i, $i \in \phi$, given the transformed item parameters $\hat{\omega}_{(1,j)i}$, $j \in \psi_i$, will be close to each other. We can have the relationship:

$$S(\hat{\omega}_{(1,j)i}) \approx S(\hat{\omega}_{(1,k)i}), \text{ for } j \in \psi_i,\ k \in \psi_i \text{ and } j \neq k. \tag{16}$$

Let $\mathbf{T}$ denote a vector whose elements are $A_{(1,j)}$ and $B_{(1,j)}$, j=2 to m. The loss function $Q(\mathbf{T})$ can be defined as

$$Q(\mathbf{T}) = \sum_{i \in \phi} \sum_{k \in \psi_i} \sum_{j \in \psi_i, j>k} \{S(\hat{\omega}_{(1,k)i}) - S(\hat{\omega}_{(1,j)i})\}^2 \tag{17}$$

where

$$S(\hat{\omega}_{(1,j)i}) = \int_{-\infty}^{\infty} P(\theta_1, \hat{\omega}_{(1,j)i}) g(\theta_1) d\theta_1, \tag{18}$$

and $g(\theta_1)$ is the ability pdf estimated in the calibration of Pool 1 items. Alternatively, the $g(\theta_1)$ in Equation 18 can be replaced by a uniform distribution. Equation 17 can be minimized by setting its first order partial derivatives to zero, that is

$$\frac{\partial Q(\mathbf{T})}{\partial A_{(1,j)}} = 0 \text{ and } \frac{\partial Q(\mathbf{T})}{\partial B_{(1,j)}} = 0, \text{ for } j=2 \text{ to } m. \tag{19}$$

The first order partial derivatives are derived as the follows:

$$\frac{\partial Q(\mathbf{T})}{\partial A_{(1,j)}} = 2\sum_{i\in\phi}\Big\{\sum_{k\in\psi_i,k>j}[S(\hat{\boldsymbol{\omega}}_{(1,j)it}) - S(\hat{\boldsymbol{\omega}}_{(1,k)it})]\frac{\partial S(\hat{\boldsymbol{\omega}}_{(1,j)it})}{\partial A_{(1,j)}} - \sum_{k\in\psi_i,k<j}[S(\hat{\boldsymbol{\omega}}_{(1,k)it}) - S(\hat{\boldsymbol{\omega}}_{(1,j)it})]\frac{\partial S(\hat{\boldsymbol{\omega}}_{(1,j)it})}{\partial A_{(1,j)}}\Big\}, \tag{20}$$

where

$$\frac{\partial S(\hat{\omega}_{(1,j)it})}{\partial A_{(1,j)}} = \int_{-\infty}^{\infty} \frac{-1.7(1-c_{ji})a_{ji}(\theta_1 - B_{(1,j)})e^{-1.7a_{ji}(\theta_1 - B_{(1,j)} - A_{(1,j)}b_{ji})A_{(1,j)}^{-1}}}{A_{(1,j)}^2[1+e^{-1.7a_{ji}(\theta_1 - B_{(1,j)} - A_{(1,j)}b_{ji})A_{(1,j)}^{-1}}]^2} g(\theta_1)d\theta_1, \tag{21}$$

and

$$\frac{\partial Q(\mathbf{T})}{\partial B_{(1,j)}} = 2\sum_{i\in\phi}\Big\{\sum_{k\in\psi_i,k>j}[S(\hat{\boldsymbol{\omega}}_{(1,j)it}) - S(\hat{\boldsymbol{\omega}}_{(1,k)it})]\frac{\partial S(\hat{\boldsymbol{\omega}}_{(1,j)it})}{\partial B_{(1,j)}} - \sum_{k\in\psi_i,k<j}[S(\hat{\boldsymbol{\omega}}_{(1,k)it}) - S(\hat{\boldsymbol{\omega}}_{(1,j)it})]\frac{\partial S(\hat{\boldsymbol{\omega}}_{(1,j)it})}{\partial B_{(1,j)}}\Big\}, \tag{22}$$

where

$$\frac{\partial S(\hat{\omega}_{(1,j)it})}{\partial B_{(1,j)}} = \int_{-\infty}^{\infty} \frac{-1.7(1-c_{ji})a_{ji}e^{-1.7a_{ji}(\theta_1 - B_{(1,j)} - A_{(1,j)}b_{ji})A_{(1,j)}^{-1}}}{A_{(1,j)}[1+e^{-1.7a_{ji}(\theta_1 - B_{(1,j)} - A_{(1,j)}b_{ji})A_{(1,j)}^{-1}}]^2} g(\theta_1)d\theta_1. \tag{23}$$

There are a number of well established numerical methods which can be used to minimize the loss function Q. The Davidon-Fletcher-Powell (DFP) algorithm was selected to minimize Equations 17. A computer program in the C language was developed using a set of subroutines provided in *Numerical Recipes* (Press, Teukosky, Vetterling & Flannery, 1992) to perform this algorithm. This program is available from the author.

Example One: Transformation for Two Item Pools

The data used in this example were obtained from administering two forms of a mathematics test, Form 1 and Form 2, to two nonequivalent groups of examinees. (Group I consisted of 1637 examinees and Group J consisted of 1636 examinees). Both forms of the test consisted of 36 multiple choice items drawn from two item pools, Pool 1 and Pool 2, 11 of which were common items. These data were used in the examples in Kolen and Brennan (1995) and are available from the authors. The ability and item parameters were calibrated using the BILOG (Mislevy & Bock, 1991).

The common scale transformation was performed over the 11 common items using the proposed method. A uniform distribution in an interval between -4 and 4 was used as the θ distribution to compute the EPC scores. This interval was the same as the one BILOG used to calibrate the item parameters. The item parameter estimates and the EPC scores on the original scales are presented in Table 1. The c -parameters estimates are not presented because they are not affected by the scale transformation. It can be seen that the EPC scores for the common items in the two pools are quite different. For example, for item 10, the EPC score is 0.369 given the parameter estimates from Pool 1 and 0.412 given the parameter estimates from Pool 2. This difference is mainly caused by the difference between the two scales.

The transformed parameters estimates were computed for the common scale (scale 1) by the proposed common scale transformation method. The transformed a- and b-parameter estimates and the original c-parameter estimates were used to compute the EPC scores. It can be seen from Table 1 that the EPC scores for the common items are very close for the two item pools. The small differences can be attributed to the errors involved in the process of parameter calibration and scale transformation.

Example Two: Transformation for Three Item Pools

The item parameter estimates of the 11 common items for Pool 1 used in Example One were used in this example as true parameters. These item parameters were linearly transformed to three

different scales representing scales for three independently calibrated item pools: Pools 1, 2 and 3. These parameters are presented in Table 2 under the label "On Original Scales". The EPC scores for the common items on the original scales are also presented in Table 2. Because these three scales were different, the EPC scores for the same common item are not the same for the three from scales.

The common scale transformation was performed in the same way as in Example One. The transformed parameters and EPC scores are presented in Table 2 under the label "On Common Scale". Because the parameters on the three from scales were linear transformations of the true parameters, no parameter calibration errors were involved. After the common scale transformation, the parameters were placed on a common scale, the *a* and *b* parameters and the EPC scores for the common items in all three item pools became identical.

Conclusion

The proposed common scale transformation method provides an alternative approach for solving the problem of incompatible IRT parameter estimates calibrated with two or more independent samples of subjects. It is a useful method for finding a common scales for item parameters from multiple item pools calibrated independently.

References

Baker, F., & Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement, 28,* 147-162.

Haebara, T. (1980). Equating logistic ability scales by a weighted least square method. *Japanese Psychological Research, 22,* 144-149.

Kim, S.; & Cohen, A. S. (1992). Effect of linking methods on detection of DIF. *Journal of Educational Measurement., 29,* 51-66.

Kolen, M.J., & Brennan, R.J. (1995) *Test equating: methods and practices.* New York: Springer-Verlag.

Lautenschlager, G. J., & Park, D.-G.(1988). IRT item bias detection procedures: issues of model misspecification, robustness, and parameter linking. *Applied Psychological Measurement, 12,* 365-376.

Lord, F.M. (1980). *Application of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Loyd, B.H., & Hoover, H.D. (1980). Vertical equating using the Rasch model. *Journal of*

Educational Measurement, 17, 179-193.

Marco, G.L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14,* 139-160.

Mislevy, R.J., & Bock, R.D. (1990). *BILOG 3* (2nd edition). IN, Mooresville: Scientific Software Inc.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P.(1992). *Numerical Recipes in C, 2nd edition.* Cambridge MA: University Press.

Stocking, M.L, & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement,* 7, 201-210.

Table 1. The Untransformed and Transformed Item Parameter Estimates and EPC Scores For Example One

		On Original Scales			On Common Scale		
Item ID	Pool ID	a	b	EPC	a	b	EPC
1	1	0.455	-0.710	0.666	0.455	-0.710	0.666
1	2	0.442	-1.335	0.699	0.384	-1.341	0.695
2	1	0.584	-0.857	0.680	0.584	-0.857	0.680
2	2	0.573	-1.321	0.717	0.498	-1.325	0.715
3	1	0.754	0.021	0.578	0.754	0.021	0.578
3	2	0.599	-0.710	0.631	0.521	-0.623	0.621
4	1	0.663	0.051	0.557	0.663	0.051	0.557
4	2	0.604	-0.354	0.578	0.525	-0.214	0.563
5	1	1.069	0.961	0.569	1.069	0.961	0.569
5	2	0.990	0.532	0.607	0.861	0.805	0.585
6	1	0.967	0.195	0.505	0.967	0.195	0.505
6	2	0.808	-0.116	0.545	0.703	0.060	0.526
7	1	0.348	2.277	0.388	0.348	2.277	0.388
7	2	0.414	2.554	0.424	0.360	3.129	0.398
8	1	1.458	1.024	0.531	1.458	1.024	0.531
8	2	1.355	0.581	0.559	1.179	0.861	0.533
9	1	0.702	2.240	0.307	0.702	2.240	0.307
9	2	0.634	1.896	0.340	0.551	2.373	0.299
10	1	1.408	1.556	0.369	1.408	1.556	0.369
10	2	1.135	1.079	0.412	0.987	1.434	0.373
11	1	1.299	2.159	0.325	1.299	2.159	0.325
11	2	0.926	2.134	0.344	0.805	2.646	0.297

Table 2. The Untransformed and Transformed Item Parameter and EPC Scores For Example Two

		On Original Scales			On Common Scale		
Item ID	Pool ID	a	b	EPC	a	b	EPC
1	1	0.501	-0.736	0.670	0.501	-0.736	0.670
1	2	0.523	-0.748	0.671	0.501	-0.736	0.670
1	3	0.546	-0.800	0.676	0.501	-0.736	0.670
2	1	0.642	-0.870	0.682	0.642	-0.870	0.682
2	2	0.671	-0.875	0.683	0.642	-0.870	0.682
2	3	0.701	-0.922	0.688	0.642	-0.870	0.682
3	1	0.830	-0.072	0.587	0.830	-0.072	0.587
3	2	0.868	-0.112	0.591	0.830	-0.072	0.587
3	3	0.905	-0.191	0.599	0.830	-0.072	0.587
4	1	0.730	-0.045	0.567	0.730	-0.045	0.567
4	2	0.763	-0.086	0.571	0.730	-0.045	0.567
4	3	0.796	-0.166	0.579	0.730	-0.045	0.567
5	1	1.176	0.783	0.584	1.176	0.783	0.584
5	2	1.229	0.705	0.591	1.176	0.783	0.584
5	3	1.283	0.593	0.600	1.176	0.783	0.584
6	1	1.064	0.086	0.517	1.064	0.086	0.517
6	2	1.112	0.039	0.522	1.064	0.086	0.517
6	3	1.161	-0.046	0.532	1.064	0.086	0.517
7	1	0.383	1.979	0.404	0.383	1.979	0.404
7	2	0.400	1.849	0.412	0.383	1.979	0.404
7	3	0.417	1.689	0.424	0.383	1.979	0.404
8	1	1.604	0.840	0.547	1.604	0.840	0.547
8	2	1.677	0.760	0.555	1.604	0.840	0.547
8	3	1.750	0.645	0.565	1.604	0.840	0.547
9	1	0.772	1.946	0.335	0.772	1.946	0.335
9	2	0.807	1.818	0.348	0.772	1.946	0.335
9	3	0.842	1.658	0.364	0.772	1.946	0.335
10	1	1.549	1.323	0.395	1.549	1.323	0.395
10	2	1.619	1.222	0.406	1.549	1.323	0.395
10	3	1.690	1.088	0.420	1.549	1.323	0.395
11	1	1.429	1.872	0.356	1.429	1.872	0.356
11	2	1.494	1.747	0.369	1.429	1.872	0.356
11	3	1.559	1.591	0.385	1.429	1.872	0.356